1º v.a $X \rightarrow N(\mu_X ; \sigma_X^2)$

m.a.s con $n=5$ $\quad X_1 ; X_2 ; X_3 ; X_4 ; X_5$ son 5 VA iID

$$T_1 = \frac{X_1+X_2+X_3+X_4+X_5}{5}$$

¿Es insesgado? : Si lo es $E[T_1]=\mu_X$

$$E[T_1] = E\left[\frac{X_1+X_2+X_3+X_4+X_5}{5}\right] = \frac{1}{5} E[X_1+X_2+X_3+X_4+X_5] =$$

$$= \frac{1}{5}\left\{E(X_1)+E(X_2)+E(X_3)+E(X_4)+E(X_5)\right\} =$$

$$= \frac{1}{5} \cdot 5\mu_X = \mu_X$$ luego T_1 es INSESGADO

$$T_2 = \frac{X_1+X_2+2X_3+X_4+X_5}{6}$$

¿Es insesgado? : Si lo es : $E[T_2]=\mu_X$.

$$E[T_2] = E\left[\frac{X_1+X_2+2X_3+X_4+X_5}{6}\right] = \frac{1}{6} E[X_1+X_2+2X_3+X_4+X_5]$$

$$= \frac{1}{6}\left\{E(X_1)+E(X_2)+2E(X_3)+E(X_4)+E(X_5)\right\}$$

$$= \frac{1}{6} \cdot 6\mu_X = \mu_X$$ luego T_2 es INSESGADO

¿Cual de los dos es más eficiente?

$$Var(T_1) = Var\left(\frac{X_1+X_2+X_3+X_4+X_5}{5}\right) = \frac{1}{25} Var[X_1+X_2+X_3+X_4+X_5]$$

$$= \frac{1}{25}\left\{Var(X_1)+Var(X_2)+Var(X_3)+Var(X_4)+Var(X_5)\right\}$$

$$= \frac{1}{25}\left[\sigma_X^2+\sigma_X^2+\sigma_X^2+\sigma_X^2+\sigma_X^2\right] = \frac{5\sigma_X^2}{25} = \frac{\sigma_X^2}{5}$$

$$Var(T_2) = Var\left(\frac{X_1+X_2+2X_3+X_4+X_5}{6}\right) = \frac{1}{36} Var(X_1+X_2+2X_3+X_4+X_5)$$

$$= \frac{1}{36}\left\{Var(X_1)+Var(X_2)+4Var(X_3)+Var(X_4)+Var(X_5)\right\}$$

$$= \frac{1}{36} 8\sigma_X^2 = \frac{2}{9}\sigma^2.$$

Como es más eficiente el que tiene menor varianza; T_1 es más

eficiente que T_2 puesto $\frac{\sigma_x^2}{5} < \frac{2}{9}\sigma_x^2$

i: No hubiera hecho falta calcular la var (T_1), puesto que $Var(\bar{X}) = \frac{\sigma_x^2}{n=5}$ y $\bar{X}$ es el estimador más eficiente de la media poblacional.

---

2. apartado A: v. a X : "tensión de ruptura (en libras) de una unidad de fibra sintética"

$X \sim N(\mu_X = ?\ ;\ \sigma_X^2 = 0'45^2 \text{ libras}^2)$

Población : $N_X \to \infty$

m. s. r

$n_X = 16$

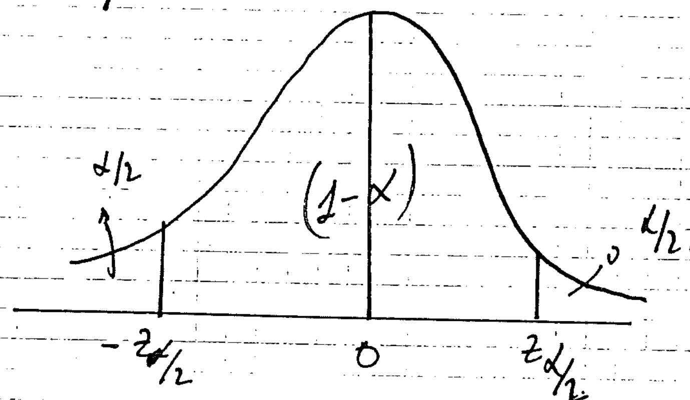
$\bar{X} = 20'3813$

v. a $\bar{X}_{16}$: "tensión media de ruptura (en libras) en una muestra aleatoria de tamaño 16".

$\bar{X}_{16} \to N\left(\mu = ?\ ;\ \sigma_{\bar{x}}^2 = \frac{(0'45)^2}{16}\right)$

Al tipificar: $\frac{\bar{X}_{16} - \mu_X}{\frac{0'45}{4}} = Z$

Gráficamente :



$P[-z_{\alpha/2} < Z < z_{\alpha/2}] = (1-\alpha)$

$P\left[-z_{\alpha/2} < \frac{\bar{X}_{16} - \mu_X}{0'45/4} < z_{\alpha/2}\right] = (1-\alpha)$

$P\left[\bar{X}_{16} - z_{\alpha/2} \cdot \frac{0'45}{4} < \mu_X < \bar{X}_{16} + z_{\alpha/2} \cdot \frac{0'45}{4}\right] = (1-\alpha)$

Si $(1-\alpha) = 0'90 \Rightarrow \alpha/2 = 0'05$

mirando en las tablas de la $N(0;1)$ $Z_{0'05} = 1'645$

La puntuación $\bar{X}_{16}$ la obtenemos sacando la media de la muestra

Por lo tanto:

$$P\left[20'3813 - 1'645 \cdot \frac{0'45}{4} \leq \mu_x \leq 20'3813 + 1'645 \cdot \frac{0'45}{4}\right] = 0'9$$

$$P\left[20'1962 \leq \mu_x \leq 20'5664\right] = 0'9 \qquad d = 0'18611$$

CONCLUSION: No podemos asegurar que la "tension promedio de ruptura" se encuentre entre 20'1962 y 20'5664, pero si podemos afirmar en un 90% de las ocasiones, que el máximo error que podemos cometer al estimar la tensión media de ruptura de todas las fibras en base a la tensión media de ruptura de una muestra de 16 fibras es de 0'1861

Para $(1-\alpha) = 0'99 \Rightarrow \alpha = 0'01 \Rightarrow \alpha/2 = 0'005 \Rightarrow Z_{\alpha/2} = 2'575$

$$P\left[20'3813 - 2'575\sqrt{\frac{0'45^2}{16}} \leq \mu_x \leq 20'3813 + 2'575 \cdot \frac{0'45}{4}\right] = 0'99$$

$$P\left[20'3813 - 0'2897 \leq \mu_x \leq 20'3813 + 0'2897\right] = 0'99$$

$$P\left[20'0916 \leq \mu_x \leq 20'6709\right] = 0'99$$

La conclusión que obtenemos, la podemos expresar en los mismos términos que antes.
Si comparamos los dos resultados, nos damos cuenta de que al incrementar $(1-\alpha)$ el intervalo es de mayor amplitud, o lo que es lo mismo que el error muestral ha aumentado.

b) Ahora σ_X^2 = desconocida, la tendremos que estimar con S^2

$$\sigma_X^2 = ? \Rightarrow \hat{\sigma}_X^2 = S^2$$

MSR pero como $N \to \infty$ MCR.

m.a.s : $n = 16$

v.a $\overline{X}_{16}$ =: "tensión media de ruptura (en libras) en una muestra de 16"

$Nat(\overline{X}_{16})$ = Normal.

$E(\overline{X}_{16}) = \mu_X$

$$Var(\overline{X}_{16}) = \frac{\sigma^2_X}{16} = ? \Rightarrow \hat{Var}(\overline{X}_{16}) = \frac{S^2}{16}$$

$$\text{Intervalo} = \overline{X}_{16} \pm k \cdot \sigma^2_{\overline{X}_{16}}$$

Intervalo = Estimador puntual ± K. desviación estandar del estimador.

?

Sustituimos por.

Estimación de la desv. estandar

$$P\left[\mu_X \in \overline{X}_{16} \pm t_{15;\alpha/2} \cdot \frac{S}{4}\right] = 1-\alpha$$

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} \Rightarrow S = 0'523092 \text{ (Con los datos de la muestra)}$$

$1-\alpha = 0'90 \Rightarrow \alpha/2 = 0'05 \Rightarrow t_{15;0'05} = 1'7531$

Luego: $P\left[\mu_X \in 20'3813 \pm 1'7531 \cdot \frac{0'523092}{4}\right] = 0'90$

$P\left[20'1520 \leq \mu_X \leq 20'6105\right] = 0'90$

El máximo error a cometer es 0'2292. Esto lo podemos afirmar con una seguridad del 90%

Si comparamos con lo obtenido en el apartado anterior.

σ_X^2 conocida. $P(20'1962 \leq \mu_X \leq 20'5664) = 0'90$

σ_X^2 ? $P(20'1520 \leq \mu_X \leq 20'6105) = 0'90$

intervalo mayor. El error máximo permitido es mayor cuando la σ^2 es desconocida

Para $1-\alpha = 0'77$

$$\bar{X} \pm t_{15;\,0'005} \frac{S}{\sqrt{n}} \Rightarrow 20'38125 \pm 2'9467 \cdot \frac{0'5231}{\sqrt{16}}$$

$$P\left[19'9959 \leq \mu_X \leq 20'77\right] = 0'99$$

El error máximo que podemos cometer es de $0'38535$, esto lo podemos afirmar con un riesgo de equivocarnos del 1%

Si lo comparamos con lo obtenido en el apartado anterior.

σ_X^2 CONOCIDA $\Rightarrow$ error máx : $0'2897$. El intervalo de confianza depende de
- dispersión
- tamaño de la muestra
- nivel de confianza.

σ_X^2 ? $\Rightarrow$ " : $0'38535$

intervalo más amplio.

Cuando la varianza es conocida, podemos afinar más en la afirmación (el intervalo será menor)

El valor de "t" siempre será mayor que el valor de "z" para un mismo nivel de confianza.
Cuando $n > 100$, la distribución "t" se transforma en una distribución "z"

3°

v.a X : "Cantidad que gasta al día una persona"

$$\mu_X = ? \quad \sigma_X^2 = ?$$

m.a.s : $n = 16$

MSR; aunque $N = 5000$ (Población finita) como $n = 16$ (No suficientemente grande) MCR.

$\bar{X}_{16}$: "Cantidad media que gastan al día una ~~media~~ muestra de 16 personas"

$\bar{X} = 158'5$
$S^2 = 16'41$ (Con los datos de la muestra)

$$P\left[\mu_X \in \bar{X} \pm t_{n-1;\,\alpha/2} \sqrt{\frac{S^2}{n}}\right] = 0'9$$

Para $(1-\alpha) = 0'9$ $\quad t_{15;\,0'05} = 1'7531$

$$P\left[\mu_X \in 158'5 \pm 1'7531 \sqrt{\frac{16'41}{16}}\right] = 0'9$$

$$P\left[151'3057 \leq \mu_X \leq 165'6943\right] = 0'9$$

Interpretación: Podemos decir que en el 90% de las ocasiones la cantidad media que gasta al día una persona que asiste a convenciones de esa ciudad estará entre $151'3057$ u.m. y $165'6943$ u.m.

NOTA: Si lo que se quiere saber es la CANTIDAD TOTAL que gastan

al día los 5000 asistentes, solo es cuestión de multiplicar esas cifras por 5000.

Para $1-\alpha = 0'95$

$$P\left[\mu_x \in 158'5 \pm 2'1353 \sqrt{\frac{16'41}{16}}\right] = 0'95$$

pues $t_{15;\, 0'025} = 2'1353$.

Entonces: $P\left[149'753 \leq \mu_x \leq 167'247\right] = 0'95$

Para $(1-\alpha) = 0'99$

$$P\left[\mu_x \in 158'5 \pm 2'9467 \sqrt{\frac{16'41}{4}}\right] = 0'99$$

Pues $t_{15,\, 0'005} = 2'9467$.

Entonces: $P\left[146'408 \leq \mu_x \leq 170'592\right] = 0'99$

Si comparamos los resultados obtenidos, podemos ver que si mantenemos constante todo lo demás, a medida que aumentamos el nivel de confianza, los intervalos son cada vez mayores.

4º

v.a X : "nº de millas recorridas durante una semana por un camión"

Nat. X : ?

μ_x = ?

$\sigma_x^2 = ? \implies \hat{\sigma}_x = 120$ millas

$d_{max} = 15$ millas

$1-\alpha = 0'9545$

$N = 500.000$

a) ¿n?

Como $N = 500000$ (finita) el MSR.

$$n_{MCR} = z^2_{\alpha/2} \frac{\sigma_x^2}{d^2_{max}} = (-2)^2 \frac{120^2}{15^2} = 256$$

Porque $z_{0'0227} = -2$ (en tablas)

$$n_{MSR} = \frac{256}{1 + \frac{256-1}{500.000}} = 255'86 \simeq \underline{\underline{256}}$$

Nota: He supuesto que el tamaño de n sería lo suficientemente grande para que por el TCL la distribución de $\bar{x}$ sea Normal.

b) Si $\sigma_x = 150$ millas

$\overline{X}_{256}$: "nº medio de millas recorridas durante una semana por un camión en una muestra de 256 camiones".

Dos son los requisitos que debería cumplir:

$$d_{max} = 15 \text{ millas}$$

$$1-\alpha = 0'9545$$

d_{max} es el radio de nuestro intervalo : $\overline{x} \pm z_{\alpha/2} \frac{\sigma_x}{\sqrt{n}}$

En nuestro caso : $2 \cdot \frac{150}{\sqrt{256}} = 18'75 > 15$ NO CUMPLE.

¿Cual tendría que ser n? :

$$n = (-2)^2 \cdot \frac{150^2}{15^2} = 400$$

Si $\sigma_x = 96$ millas

Por el mismo razonamiento que antes : $2 \cdot \frac{96}{\sqrt{256}} = 12 < 15$ CUMPLE

En este caso bastaría con un tamaño de $n = (-2)^2 \cdot \frac{96^2}{15^2} = 163'84 \simeq 164$

c) Si $N = 2.000.000$ camiones.

Veamos que pasa en los tres casos presentados.

$$n = \frac{256}{1 + \frac{255}{2.000.000}} = 255'96 \simeq 256$$

$$n = \frac{400}{1 + \frac{399}{2.000.000}} = 399'92 \simeq 400$$

$$n = \frac{164}{1 + \frac{163}{2.000.000}} = 163'98 \simeq 164$$

NO VARÍA EL TAMAÑO DE n.

5 v.a X: "Monto medio de activos financiados por creditos a largo plazo de una empresa"

$N = 500$

$1-\alpha = 0'9$

$d_{max} = 50.000$.

$[100.000 < \hat{\sigma}_x < 500.000]$

Como la desviación tipica toma cualquier valor entre esos dos, vamos a coger los valores extremos, y en función de eso decidiremos.

Para $\hat{\sigma}_x = 100.000$

MCR $n = (Z_{\alpha/2})^2 \dfrac{\sigma_x^2}{d_{max}^2} = (-1'645)^2 \dfrac{(100.000)^2}{(50.000)^2} = 10'8241 \simeq 11$

Como un valor de $Z_{0'05} = 1'645$ el valor de $Z_{0'95}$ será $-1'645$

MSR $n = \dfrac{11}{1+\frac{10}{500}} = 10'78 \simeq 11$

Para $\hat{\sigma}_x = 500.000$

MCR $n = (-1'645)^2 \dfrac{(500.000)^2}{(50.000)^2} = 270'6025 \simeq 271$

MSR $n = \dfrac{271}{1+\frac{270}{500}} = 175'97 \simeq 176$

Por lo tanto con un tamaño de n de 176 sería suficiente para los propositos del analista.

---

6 v.a X: "Peso (en kg) de una maleta de un pasajero en un vuelo de avión"

Nat X: ? ; $\mu_x = ?$ $\sigma_x^2 = ?$

$N \to \infty$ MSR (como $N \to \infty$) $\Rightarrow$ MCR

m.a.s. : $n = 50$ en donde $\bar{X} = 27'8$ kg. $S = 5'8$ kg.

v.a $\bar{X}_{50}$: "Peso medio (en kg) de una maleta en una muestra de 50 maletas"

v.a. $\bar{X}_{50} \xrightarrow{TCL} N\left(\mu_{\bar{X}} = \mu_x = ? \; ; \; \sigma_{\bar{X}}^2 = \dfrac{\sigma_x^2}{\sqrt{50}} = ?\right)$

Nos piden el error máximo que podemos cometer para un $(1-\alpha) = 0'9$.

Como no sabemos σ_X^2 tendremos que estimarla en la muestra $\hat{\sigma}_X^2 = S^2 =$
$= 5'8^2$ Kg.² ... Entonces:

$$v.a\ \overline{X}_{50} \longrightarrow t_{49}$$

Como sabemos que $P\left[\mu_X \in \overline{X} \pm t_{49;0'05} \frac{S}{\sqrt{n}}\right] = 0'9$.

lo que nos piden es $d\,max = t_{49;0'05} \frac{S}{\sqrt{n}} = 1'6766 \cdot \frac{5'8}{\sqrt{50}} =$

$= 1'3752$

Con una probabilidad de 0'9 (en el 90% de las ocasiones) podemos decir que el máximo error que podemos cometer en la estimación de μ_X es de $1'3752$ Kg.

7° v.a X : "Peso medio en onzas de un bote de piña"

Nat $X = ?$ $\mu_X = ?$ $\sigma_X^2 = ?$

m.a.s : $n = 50$ en donde $\overline{X} = 67'2$ y $S = 2'1$

nos piden ¿$P[67'2 - 0'4 \leq \overline{X}_{50} \leq 67'2 + 0'4]$?

¿$\overline{X}_{50}$? $\longrightarrow$ v.a $\overline{X}_{50} \xrightarrow{TCL} N\left(\mu_{\overline{X}} = \mu_X = ?\ ;\ \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{\sqrt{n}} = ?\right)$

como no conocemos σ_X^2 lo estimamos por $S^2 = \hat{\sigma}_X^2$ entonces:

$$\overline{X}_{50} \longrightarrow t_{49}$$

$$P[67'2 - 0'4 \leq \overline{X}_{50} \leq 67'2 + 0'4] = P\left[\overline{X} \pm t_{49;\alpha/2} \frac{S}{\sqrt{50}}\right] =$$

$$= P\left[\overline{X} \pm t_{49,\alpha/2} \cdot \frac{2'1}{\sqrt{50}}\right]$$

Lo que tenemos que encontrar es el valor de $(1-\alpha)$.

Sabemos que: $t_{49,\alpha/2} \cdot \frac{2'1}{\sqrt{50}} = 0'4$

Luego: $t_{49;\alpha/2} = \frac{0'4 \cdot \sqrt{50}}{2'1} = 1'3468$

Este valor será el que deje a su derecha un área de $\alpha/2$ bajo la curva de una t_{49} ~~fac~~.

Mirando en tablas :

Δt	$\Delta \alpha$
$1'2991 - 1'6766 = 0'3775$	$-0'05$
$1'2991 - 1'3468 = 0'0477$	x

$$x = \frac{-0'05 \cdot 0'0477}{0'3775} = 0'0063$$

Luego α $0'10 - 0'0063 = 0'0937 = \alpha/2$

Nos piden $(1-\alpha) = 1 - (0'0937) \cdot 2 = \underline{0'8126}$

<u>Nota</u>: Aunque el MSR como $N \to \infty$ he considerado el MCR.

8º

v.a. X: "Numero de defectos en una pieza del pedido.
$Nat\ X = ?$ $\mu_X = ?$ $\sigma_X^2 = ?$

m.a.s.: $n = ?$

MSR pero como $N \to \infty \Rightarrow$ MCR

$d_{max} = $ (Parametro − estimador) $= 0'2$ def/pieza

$(1-\alpha) = 0'9545$

$\hat{\sigma}_X^2 = (1'2)^2$ (def/pieza)^2.

a) $$n_{MCR} = Z^2_{\alpha/2} \frac{\sigma_X^2}{d^2_{max}} = 2^2 \cdot \frac{(1'2)^2}{(0'2)^2} = 144$$

$Z_{0'02275}$ = mirando en tablas $= 2$ (Valor de Z que me deja un area de $0'0227$ a la derecha).

Con $n = 144$ garantizamos lo requerido.

b) Con $n = 144$ m.a.s. donde obtenemos $\bar{X} = 2$; $S = 1'6$

$\bar{X}_{144}$: "nº medio de defectos/pieza en una muestra de 144 piezas"

$Nat\ \bar{X}_{144}$ = Normal (TCL)

$E[\bar{X}_{144}] = \mu_X = ?$

$Var(\bar{X}_{144}) = \dfrac{\sigma_X^2}{144} = ?$

Como desconocemos σ_X^2 sustituimos por su estimación en la muestra S^2

Entonces $\bar{X}_{144} \to t_{143}$

~~Para~~

Por lo tanto: $P\left[\mu_X \in \bar{x} \pm t_{143} \dfrac{S}{\sqrt{144}}\right] = 0'9545$

Como los grados de libertad son elevados podemos perfectamente considerar que $\bar{X}_{144} \to$ Normal, por lo que:

$$P\left[\mu_X \in \bar{x} \pm Z_{0'02275} \cdot \frac{1'6}{\sqrt{144}}\right] = 0'9545$$

$$P\left[\mu_X \in 2 \pm 2 \cdot \frac{1'6}{\sqrt{144}}\right] = 0'9545$$

2 será el valor de Z que nos deje a su derecha un area de $0'02275$.

$P[\mu_x \in 2 \pm 0'26] = 0'9545$

El máximo error que podemos cometer al estimar al estimar el número medio de defectos/pieza en base al promedio obtenido en una muestra de 144 piezas es de $0'2\hat{6}$ con una probabilidad de $0'9545$ (en el 95'45% de las ocasiones)

c) No, porque el error muestral máximo que hemos obtenido ($0'2\hat{6}$) es superior al error muestral máximo que pretendíamos ($0'2$)

d) La varianza estaba mal estimada, pues se estimó en $(1'2)^2$ y luego a resultado $(1'6)^2$, es decir, estaba infravalorada la estimación de la varianza.

e) Nuestro intervalo será: $P[\mu_x \in 2 \pm 0'2\hat{6}] = 0'9545$

es decir, aproximadamente entre $1'73$ y $2'27$, y la probabilidad de que la verdadera media esté ahí (entre esos valores) es de $0'9545$. Dado que el valor 1 no está dentro de ese intervalo, y por lo tanto ahora habrá más defectos por pieza que antes y por lo tanto la calidad es menor.

9º v.a X: "nº de veces que se enciende la dimmer a la semana en una casa"

$Nat(X) = ?$ $\mu_x = ?$ $\sigma_x^2 = ?$

$N = 20.000$ m.a.s. : $n = 400$ en donde se obtienen que $\bar{x} = 4$ veces/semana $S = 0'6$ veces/semana

¿ intervalo de confianza para $(1-\alpha) = 0'95$?

$\bar{X}_{400}$: "nº medio de veces por semana que una casa enciende la dimmer en una muestra de 400 casas

$Nat.(\bar{X}_{400}) \xrightarrow{TCL}$ Normal

$\mu_{\bar{X}_{400}} = \mu_x = ?$

$\sigma^2_{\bar{X}_{400}} = \frac{\sigma_x^2}{400} = ? \Rightarrow \hat{\sigma}_x^2 = S^2 = (0'6)^2$

Como no conocemos la varianza poblacional, la sustituimos por su estimación en la muestra y entonces ya no estaríamos ante una distribución NORMAL, sino ante una t de Student con 399 g.l. Como los g. de l. son elevados (elevado tamaño

de la muestra) estaríamos otra vez ante una distribución Normal.

M.S.R. pero como N es grande ($\to \infty$) M.C.R.

intervalo: $\bar{X} \pm Z_{\alpha/2} \frac{S_x}{\sqrt{n}}$

$(1-\alpha) = 0'95 \Rightarrow \alpha/2 = 0'025 \Rightarrow Z_{0'025} = 1'96$ (mirando tablas)

Luego: $P\left[\mu_x \in 4 \pm 1'96 \frac{0'6}{\sqrt{400}}\right] = 0'95$

$P[\mu_x \in 4 \pm 0'0588] = 0'95$

$P[3'9412 \leq \mu_x \leq 4'0588] = 0'95$

En el 95% de las ocasiones podemos decir que el nº medio de veces que se enciende una chimenea a la semana oscilará entre 3'9412 y 4'0588. El error máximo que podemos cometer será de 0'0588 veces a la semana.

Si se consumen 5 Kg cada vez que se enciende la chimenea, Entonces.

$P[19'706 \leq \mu_y \leq 20'294] = 0'95$ donde v.a Y: "Kg que se consumen en una casa con chimenea".

10°

v.a X: "Ventas con campaña en una tienda"

v.a Y: "Ventas sin campaña en una tienda"

$d = X - Y$ v.a d: "Diferencia de ventas con y sin campaña en una tienda"

M.a.S.: $n_x = 13$ $n_y = 13$ Dependientes pues la 2ª muestra no se elige aleatoriamente.

diferencias en la muestra.

Tienda	dif. (x-y)
1	1'9
2	4'7
3	-1'2
4	7'8
5	5'3
6	0'5
7	4'9
8	5'3
9	10'7
10	5'9
11	-1'3
12	7'6

$\bar{d} = 4'038$

$S_d = 3'71$

v.a X   Nat: Suponemos Normal
μ_x : ?
σ_x^2 : ?

v.a Y   Nat: Suponemos Normal
μ_y : ?
σ_y^2 : ?

v.a $D = X - Y$: Nat: Normal

$$E[D] = E[X-Y] = \mu_x - \mu_y = ?$$
$$Var[D] = Var(X-Y) = Var(X) + Var(Y) = ?$$

v.a $\bar{d}_{13} = \frac{d_1 + d_2 + \cdots + d_{13}}{13}$: "Diferencia media de ventas antes y despues de la promoción en una muestra de 13 tiendas"

Nat $(\bar{d})$: Normal

$E(\bar{d}) = \mu_d = \mu_x - \mu_y = ?$

$Var(\bar{d}) = \frac{\sigma^2_d}{n} = ?$ $\hat{Var}[\bar{d}] = \frac{S^2_d}{n}$

MSR pero como $N \to \infty \Rightarrow$ MCR.

Para $(1-\alpha) = 0'95$ $P[\mu_d \in \bar{d} \pm t_{12;0'025} \cdot \frac{S_d}{\sqrt{n}}] = 0'95$

$$P[\mu_d \in 4'038 \pm 2'1788 \frac{3'71}{\sqrt{13}}] = 0'95$$

$$P[1'796 \leq \mu_d \leq 6'2799] = 0'95$$

El intervalo es positivo, por lo que las ventas han aumentado con la promoción. Nos hemos podido equivocar con una probabilidad de 0'05.

---

11.   v.a X : "Numero de accidentes en una planta durante un mes antes de aplicar el programa de seguridad"

Nat X : Suponemos NORMAL $\mu_x = ?$ $\sigma_x^2 = ?$

v.a Y : "Numero de accidentes en una planta durante un mes despues de aplicar el programa de seguridad"

Nat Y : Suponemos Normal. $\mu_y = ?$ $\sigma_y^2 = ?$

v.a $D = X - Y$: "Diferencia de nº de accidentes en una planta/mes antes y despues de aplicar el programa de seguridad.

Nat D : Normal $\mu_d = \mu_x - \mu_y = ?$ $\sigma_d^2 = \sigma_x^2 + \sigma_y^2 = ?$

m.a.s : $n_x = 6$ $n_y = 6$ donde: $\bar{d} = 4$; $S_d^2 = 10'4$

M S R $\Rightarrow$ $N \to \infty$ $\Rightarrow$ MCR

v.a $\bar{d} = \dfrac{d_1 + \cdots + d_6}{6}$: "Dif. media de accidentes en una muestra de 6 observaciones antes y despues de aplicar el programa de seguridad."

Nat $\bar{d}$: Normal

$E[\bar{d}] = \mu_d = \mu_x - \mu_y = ?$

$Var(\bar{d}) = \dfrac{\sigma^2 d}{n} = ?$ $\hat{Var}(\bar{d}) = \dfrac{S^2 d}{n}$

v.a $\bar{d} \to t_{n-1}$

Para $(1-\alpha) = 0'90$ $P\left[\mu_d \in \bar{d} \pm t_{n-1;\,0'05} \sqrt{\dfrac{S_d^2}{n}}\right] = 0'90$

$P\left[\mu_d \in 4 \pm 2'015 \sqrt{\dfrac{10'4}{6}}\right] = 0'90$

Porque el valor de una $t_{5;\,0'05} = 2'015$ (mirando tablas)

$P\left[1'3463 \le \mu_d \le 6'6537\right] = 0'90$

Podemos decir que en el 90% de las ocasiones la diferencia media de accidentes antes y despues de aplicar el programa de seguridad, estara entre 1'3463 accidentes y 6'6537 accidentes.

¿Ha sido efectivo el programa? : Como $\mu_d = \mu_x - \mu_y > 0$ Ha sido efectivo con una probabilidad de 0'90 ya que se han reducido accidentes $\mu_y < \mu_x$

---

12 v.a X : "Tiempo (en seg.) empleado por un operario para realizar la tarea siguiendo el metodo A"

Nat (x) : Supongo Normal $\mu_x = ?$ $\sigma_x^2 = ?$

v.a Y : "Tiempo (en seg) empleado por un operario para realizar la tarea siguiendo en metodo B"

Nat (Y) : Supongo Normal $\mu_y = ?$ $\sigma_y^2 = ?$

m.a.s. : $n_x = 12$ $n_y = 15$ Los operarios sobre los que se realizan ambas muestras son diferentes, luego las v.a. son independientes.

donde obtenemos que: $\bar{X} = 304$ seg.

$S_x = 18$ seg.

$\bar{Y} = 325$ seg.

MSR pero como $N \to \infty$ (en ambos casos) MCR.

v. a $\overline{X}_{12} = \frac{x_1 + \cdots + x_{12}}{12}$: "Tiempo medio (en seg.) empleado por una muestra de 12 operarios utilizando el método A"

Nat. $\overline{X}_{12}$: Suponemos Normal $\mu_{\overline{X}_{12}} = \mu_X = ?$ $\sigma^2_{\overline{X}_{12}} = \frac{\sigma_X^2}{n_X} = ?$

v. a $\overline{Y}_{15} = \frac{y_1 + \cdots + y_{12}}{15}$

Nat $\overline{Y}_{15}$: Suponemos Normal $\mu_{\overline{Y}_{15}} = \mu_Y = ?$ $\sigma^2_{\overline{Y}_{15}} : \frac{\sigma_Y^2}{n_Y} = ?$

Se suponen Normales, porque sino como n_X y $n_Y < 30$ No se podría realizar el problema.

v. a $(\overline{X}_{12} - \overline{Y}_{15})$: Nat $\to$ Normal

$\mu_{\overline{X}_{12}} - \mu_{\overline{Y}_{15}} = ?$

$Var[\overline{X}_{12} - \overline{Y}_{12}] = \frac{\sigma_X^2}{12} + \frac{\sigma_Y^2}{15} = ?$

$\hat{Var}(\overline{X}_{12} - \overline{Y}_{15}] = \frac{S_X^2}{12} + \frac{S_Y^2}{15}$.

Hay que preguntarse si las varianzas poblacionales aunque desconocidas son iguales o diferentes : <u>Prueba de Homocedasticidad</u>.

Suponemos $\sigma_X^2 = \sigma_Y^2$.

$\frac{S_X^2}{S_Y^2} = F_{n_X-1;\, n_Y-1} \Rightarrow \frac{18^2}{24^2} = 0'5625$

Valor crítico : $(1-\alpha) = 0'99$

$F_{11;14;0'005} = 4'568$

$F_{11;14;0'995} = 0'195$

El valor $0'5625$ queda entre estos dos valores, luego no tenemos pruebas suficientes para rechazar que $\sigma_X^2 = \sigma_Y^2$; Es decir, son iguales o no significativamente distintas.

Por lo tanto : $\hat{Var}[\overline{X}_{12} - \overline{Y}_{15}] = \frac{S_P^2}{12} + \frac{S_P^2}{15}$

<u>Intervalo</u> : $(\overline{X}_{12} - \overline{Y}_{15}) \pm t_{n_X+n_Y-2;\, \alpha/2} \sqrt{\frac{S_P^2}{12} + \frac{S_P^2}{15}}$

$(\overline{X}_{12} - \overline{Y}_{15}) = (304 - 325)$

$$S_p^2 = \frac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x+n_y-2} = \frac{12 \cdot 18^2 + 14 \cdot 24^2}{25} = 465'12$$

$$t_{25;\,0'005} = 2'78794$$

$$P\left[(\mu_x-\mu_y) \in -31 \pm 2'78794 \sqrt{\frac{465'12}{12} + \frac{465'12}{15}}\right] = 0'99$$

$$P\left[(\mu_x-\mu_y) \in -31 \pm 23'28\right] = 0'99$$

$$P\left[-54'28 \leq \mu_x-\mu_y \leq -7'72\right] = 0'99$$

Con una probabilidad de 0'99 de acertar, podemos decir que el máximo error que podemos cometer al estimar la diferencia entre las dos medias poblacionales ($\mu_x-\mu_y$) en base a las medias obtenidas en las muestras es de 23'28.

El Tiempo empleado por los operarios que utilizan el método B es mayor que el empleado por los operarios que utilizan el metodo A., pues el intervalo es negativo.

¿Por qué? → No hay evidencia de que $\mu_x=\mu_y$ al nivel de confianza del 99%

13

v.a X : "Nº de decibelios medidos a 15m. que emite un camion que circula como maximo a 60km/h en zona de trafico interestatal"

Nat (x) : ? μ_x = ? σ_x^2 = ?

v.a Y : "Nº de decibelios medios a 15m., que emite un autobus que circula como maximo a 60km/h en zona de trafico interestatal"

Nat (Y) : ? μ_y = ? σ_y^2 = ?

m.a.S : n_x = 50 camiones donde μ_x = 91'5 decibelios
S_x = 5'2 "

m.a S : n_y = 50 autobuses donde μ_y = 88'6 decibelios
S_y = 3'8 "

MSR pero como $N \to \infty$: MCR

Objetivo : Construir un intervalo para el parámetro ($\mu_x-\mu_y$) con $(1-\alpha)=0'95$

$(\bar{x}-\bar{y})$ - Radio

Para obtener el radio, tenemos que saber cómo se distribuye la v.a $(\bar{X}-\bar{Y})$:

v.a $(\bar{X}) \xrightarrow{TCL}$ Normal $\left(\mu_{\bar{X}}=\mu_X=?\ ;\ \sigma^2_{\bar{X}}=\frac{\sigma^2_X}{n}=?\right)$

v.a $(\bar{Y}) \xrightarrow{TCL}$ Normal $\left(\mu_{\bar{Y}}=\mu_Y=?\ ;\ \sigma^2_{\bar{Y}}=\frac{\sigma^2_Y}{n}=?\right)$

v.a $(\bar{X}-\bar{Y}) \longrightarrow$ Normal $\left(\mu_{\bar{X}}-\mu_{\bar{Y}}=\mu_X-\mu_Y=?\ ;\ \sigma^2_{\bar{X}-\bar{Y}}=\frac{\sigma^2_X}{n}+\frac{\sigma^2_Y}{n}=?\right)$

Al ser las varianzas desconocidas tendremos que sustituir por su estimación en la muestra, y entonces estamos ante una t de student con g.l. que dependerán de si las varianzas poblacionales son iguales o distintas: PRUEBA DE HOMOCEDASTICIDAD

Supongamos $\sigma^2_X=\sigma^2_Y$

Entonces $\frac{S^2_X}{S^2_Y}=F_{n_X-1;\,n_Y-1}=\frac{(5'2)^2}{(3'8)^2}\simeq 1'87$

Valor crítico: $F_{49;49}$ = No hay en tablas: Aproximamos por defecto y calcularemos $F_{40;40}$

Para $(1-\alpha)=0'95$ $\quad F_{40;40;0'025}=1'88$

$F_{40;40;0'975}=\frac{1}{F_{40;40;0'025}}=\frac{1}{1'88}=0'5319$

1'87 está dentro de ese intervalo de valores; si cogiésemos $F_{49;49}$ el límite superior sería menor luego 1'87 no entraría: No aceptamos el supuesto de $\sigma^2_X=\sigma^2_Y$

No se cumple el supuesto de Homocedasticidad y $\sigma^2_X\neq\sigma^2_Y$

Nuestro intervalo sería entonces:

$$(\bar{X}-\bar{Y})\pm t_{k;\alpha/2}\sqrt{\left(\frac{\hat{\sigma}^2_X}{n_X}=\frac{s^2_X}{n_X}\right)+\left(\frac{\hat{\sigma}^2_Y}{n_Y}=\frac{s^2_Y}{n_Y}\right)}$$

$(\bar{X}-\bar{Y})=(91'5-88'6)$

$$K=\frac{(50-1)(50-1)}{(50-1)\,c^2+(1-c)^2(50-1)} \quad \text{donde } c=\frac{\left(\frac{5'2}{50}\right)^2}{\left(\frac{5'2}{50}\right)^2+\left(\frac{3'8}{50}\right)^2}=0'8994$$

Sustituyendo c en la fórmula obtenemos $K=49'5177\simeq 50$

$$(91'5-88'6)\pm t_{50;0'025}\cdot\sqrt{\frac{5'2^2}{50}+\frac{(3'8)^2}{50}}$$

$$2'9 \pm 2'0086 \cdot 0'9108$$

pues $t_{50;\,0'025} = 2'0086$ (mirando en tablas)

$$\sqrt{\frac{5'2^2}{50} + \frac{3'8^2}{50}} = 0'9108$$

$$2'9 \pm 1'82948$$

$$P[1'0705 \leq \mu_x - \mu_y \leq 4'72948] = 0'95$$

En el 95% de las ocasiones la diferencia de decibelios emitidos ...
... por los camiones y los autobuses, oscilará entre 1'0705 y 4'72948.
Por lo tanto, podemos afirmar que los camiones emiten más decibelios que los autobuses, aunque nos podemos equivocar el 5% de las veces

---

14° v.a X : "Lectura en parte de millón de un intervalo"

$Nat(x) = ?$ $\mu_x = ?$ $\sigma_x^2 = ?$

m.a.s : $n = 6$ donde : $\bar{x} = 9'485$ $S_x = 0'1689$
(De los datos de la muestra)

<u>Objetivo</u>: Intervalo para σ_x^2 con $(1-\alpha) = 0'9$

Sabemos que $\frac{(n-1)S^2}{\sigma^2} \simeq \chi^2_{n-1}$

Por lo que : $P\left[\frac{(n-1)S^2}{\chi^2_{(n-1)\,\alpha/2}} \leq \sigma_x^2 \leq \frac{(n-1)S^2}{\chi^2_{(n-1);\,(1-\alpha/2)}}\right] = 0'9$

Sustituyendo : $P\left[\frac{5\cdot(0'1689)^2}{11'071} \leq \sigma_x^2 \leq \frac{5(0'1689)^2}{1'145}\right] = 0'9$

$$P[0'0129 \leq \sigma_x^2 \leq 0'1247] = 0'9$$

El valor de $\chi^2_{5;\,0'05} = 1'145$

$\chi^2_{5;\,0'95} = 11'071$

En el 90% de las ocasiones el verdadero valor de σ_x^2 estará entre 0'0129 y 0'1247.

Estas probabilidades esperadas bajo el supuesto de independencia (hipótesis nula cierta) que acabamos de calcular para cada celda o casilla, y que son iguales al producto de las probabilidades marginales de su correspondiente fila y columna, son la base para empezar a construir un estadístico de contraste.

1.3 Estadísticos de contraste

El estadístico de contraste y la regla o criterio de decisión son los instrumentos que nos van a permitir decidir de forma objetiva qué hacer (no rechazar o no aceptar) con la hipótesis nula de independencia entre las características analizadas en la tabla de contingencia.

Existen varias opiniones sobre los estadísticos de contraste a calcular en las tablas de contingencia. Algunos autores afirman que puede establecerse para la globalidad de la tabla un único estadístico de contraste que nos exprese la asociación o independencia entre las características. Otros autores defienden la postura de que encontrar un único indice de asociación para toda la tabla no tiene mucho sentido, y que lo más que podemos llegar a conocer es la existencia de una cierta asociación para algunas de las categorías de las características de la tabla de contingencia.

Sin dejar de lado ninguna de las opiniones, mostraremos en primer lugar una serie de estadísticos de contraste que estudian la existencia de asociación para la globalidad de la tabla, y mostraremos después algunas pinceladas de la opinión de los que consideran que sólo podemos analizar la asociación para pares de categorías.

1.3.1 Estadísticos de contraste para la globalidad de la tabla

1.3.1.1 El modelo esperado

En nuestro anterior análisis habíamos llegado a calcular las probabilidades esperadas para cada una de las intersecciones de las categorías de las características de interés, esto es, habíamos llegado a determinar las probabilidades esperadas para cada una de las casillas en el caso de que se cumpliese la hipótesis nula de independencia.

Una vez calculadas las probabilidades esperadas para cada una de las casillas, podemos obtener fácilmente **cuántas observaciones deberían haber aparecido** en cada uno de los cruces o intersecciones de las categorías en el caso de independencia entre las características (hipótesis nula cierta). Si disponemos adecuadamente estas **frecuencias esperadas** en forma de matriz, tendremos el **modelo teórico o esperado bajo la hipótesis nula de independencia**.

En otras palabras, el modelo teórico o esperado es también una tabla de contingencia, en cuyas celdas o casillas aparecen las frecuencias que esperaríamos haber encontrado en nuestro estudio si la hipótesis nula fuese cierta.

Bajo el supuesto de independencia, las frecuencias observadas del modelo teórico se obtienen multiplicando el tamaño de muestra por la probabilidad teórica de cada casilla:

$$E_{ij} = n \cdot P_{ij} = n \cdot P_{i.} \cdot P_{.j} =$$

$$= n \cdot \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} = \frac{n_{i.} \cdot n_{.j}}{n}$$

Para nuestro ejemplo del estudio sobre la opinión de los individuos acerca de la igualdad de sexos, el modelo esperado (léase, modelo de frecuencias esperadas) es:

	MUJERES	HOMBRES	
A FAVOR	247	305	552
EN CONTRA	156	192	348
	403	497	900

Evidentemente, las frecuencias marginales esperadas deben **siempre** coincidir con las frecuencias marginales observadas, ya que es a partir de éstas desde donde se obtienen las frecuencias esperadas para cada una de las casillas.

Este es un buen momento para comprobar que el modelo esperado es el que cumple a la perfección la hipótesis nula de independencia:

Por ejemplo, comprobemos que en dicho modelo esperado la probabilidad de que una persona manifieste encontrarse a favor de la igualdad de sexos es la misma para los hombres que para las mujeres, siendo además igual para la totalidad de la muestra:

$$P(A\ Favor) = P(A\ Favor/Hombre) = P(A\ Favor/Mujer)$$

$$\frac{552}{900} = \frac{305}{497} = \frac{247}{403} = 0{,}613$$

Esto se puede expresar del siguiente modo: el 61% de las personas encuestadas manifiesta encontrarse a favor de la igualdad de sexos. La independencia significa entonces que el 61% de los hombres (el mismo porcentaje "a favor" que en el total de la muestra) debería encontrarse a favor de la igualdad de sexos y que también el 61% de las mujeres debería encontrarse a favor de la igualdad de sexos.

Comprobemos también que en el modelo esperado la probabilidad de que una persona encuestada sea mujer es igual para los que están a favor y para los que están en contra de la igualdad de sexos:

$$P(Mujer) = P(Mujer/A\ Favor) = P(Mujer/En\ Contra)$$

$$\frac{403}{900} = \frac{247}{552} = \frac{156}{348} = 0{,}448$$

Esto se puede interpretar del siguiente modo: el 45% de las personas encuestadas son mujeres. La independencia significa entonces que también el 45% de las personas que manifiestan estar a favor de la igualdad de sexos deberían ser mujeres y que, igualmente, el 45% de las personas que manifiestan estar en contra de la igualdad de sexos deberían ser mujeres.

Respecto a la interpretación del significado de la independencia conviene desterrar la idea de que independencia es necesariamente lo mismo que equiprobabilidad. **INDEPENDENCIA <u>NO ES NECESARIAMENTE</u> EQUIPROBABILIDAD.** La independencia en nuestro ejemplo no significa que el 50% de las personas encuestadas estén a favor y el 50% en contra de la igualdad de sexos, ni tampoco que el 50% sean hombres y el 50% mujeres. Independencia no es tampoco que el 25% sean mujeres a favor, el 25% mujeres en contra, el 25% hombres a favor

y el 25% mujeres en contra. Esto es equiprobabilidad que no coincide, en nuestro ejemplo, con el concepto de independencia.

Independencia significa sencillamente lo siguiente: la probabilidad de que un individuo pertenezca a una determinada categoría de la característica A permanece constante para todas las categorías de la característica B, al tiempo que la probabilidad de que un sujeto pertenezca a una determinada categoría de la característica B permanece constante para todas las categorías de la característica A.

1.3.1.2 Los modelos de residuos

Una vez que disponemos de la tabla de frecuencias observadas (modelo observado) y de la tabla de frecuencias esperadas bajo la hipótesis nula (modelo teórico o esperado), es sencillo calcular las diferencias entre ellos. El **modelo de residuos** es una tabla de contingencia en la que cada casilla muestra la diferencia entre la frecuencia observada y la frecuencia esperada de dicha casilla:

$$R_{ij} = N_{ij} - E_{ij}$$

Lo ideal para no rechazar la hipótesis nula de independencia sería que todos los residuos fuesen iguales a cero, ya que esto indicaría que el modelo observado coincide plenamente con el modelo esperado bajo la hipótesis nula. Sin embargo, para no rechazar la hipótesis nula no se exige que todos los residuos sean nulos.

Para el estudio de la opinión sobre la igualdad de sexos, el modelo de residuos es el siguiente:

	MUJERES	HOMBRES	
A FAVOR	7	-7	0
EN CONTRA	-7	7	0
	0	0	0

De la lectura del modelo de residuos se desprende que en nuestro estudio hemos observado que hay:

* 7 mujeres más a favor de la igualdad de sexos de las que cabría esperar en el caso de independencia entre el sexo y la opinión.

* 7 hombres menos a favor de la igualdad de sexos de los que cabría esperar en el caso de independencia entre el sexo y la opinión.

* 7 mujeres menos en contra de la igualdad de sexos de las que cabría esperar en el caso de independencia entre el sexo y la opinión.

* 7 hombres más en contra de la igualdad de sexos de los que cabría esperar en el caso de independencia entre el sexo y la opinión.

El modelo de residuos nos mide cuánto se alejan las frecuencias observadas de las frecuencias esperadas. Es importante señalar que los residuos marginales deben ser siempre iguales a 0, ya que indican la diferencia entre las frecuencias observadas marginales y las frecuencias esperadas marginales (que como ya hemos dicho son iguales).

Sin embargo todos los residuos "brutos" no tienen la misma importancia relativa. A nadie se le escapa intuitivamente que, aunque absolutamente valgan lo mismo, no es igual la distorsión que supone un residuo 7 cuando la frecuencia esperada es 100 que cuando la frecuencia esperada es 1000. Un residuo 7 con una frecuencia esperada 100 indica una distorsión mucho más grave (es relativamente más importante, mayor) que un residuo 7 con una frecuencia esperada 1000 (es relativamente menor). **La importancia relativa del residuo de una casilla depende entonces del valor de la frecuencia esperada en dicha casilla.** De aquí surge la necesidad **de estandarizar los residuos**, es decir, de ponerlos en relación con la frecuencia esperada correspondiente a esa casilla.

Para conseguir esto, convertimos los residuos "brutos" en **residuos estandarizados**, mediante la siguiente transformación:

$$SR_{ij} = \frac{R_{ij}}{\sqrt{E_{ij}}}$$

Para nuestro estudio sobre la opinión acerca de la igualdad de sexos, los residuos estandarizados son:

	MUJERES	HOMBRES	
A FAVOR	0,434	-0,391	-----
EN CONTRA	-0,547	0,492	-----
	-----	----	-----

Vemos a partir del análisis de los residuos estandarizados que aquellos residuos que sin estandarizar eran todos iguales a 7 resultan más o menos importantes según sea la frecuencia esperada de la celda a la que corresponden. Por ejemplo, en nuestro caso, el hecho de que haya 7 mujeres menos en contra de las que cabría esperar es más relevante que el hecho de que haya 7 hombres más a favor de los que cabría esperar (cuando decimos más relevante queremos indicar que la frecuencia observada supone una desviación mayor respecto a la frecuencia esperada, o lo que es lo mismo, un residuo relativamente más importante).

A partir del momento en que definimos los residuos estandarizados, los valores marginales (que en este caso serían los residuos estandarizados marginales) dejan de tener sentido (no suman 0 ni tienen por qué sumar 0 ya que se encuentran ajustados por la frecuencia esperada de su correspondiente casilla).

1.3.1.3 Chi cuadrado de Pearson

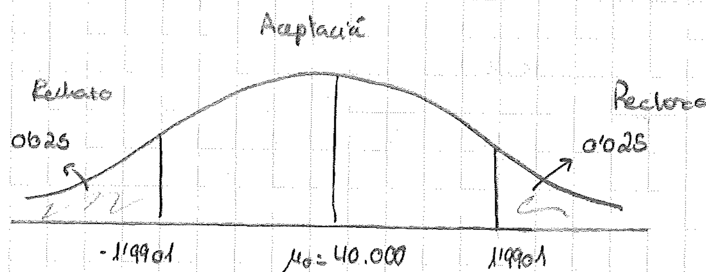
Hemos señalado con antelación el hecho de que lo ideal para no rechazar la hipótesis nula de independencia sería que el modelo de residuos contuviese en todas sus celdas un valor nulo. Esto indicaría que el modelo observado coincidiría plenamente en todas sus casillas con el modelo teórico o esperado bajo la hipótesis de independencia. Por ello el estadístico de contraste se basa en el modelo de residuos.

Sin embargo, debemos convenir que aceptar la hipótesis nula no puede exigir necesariamente que todos los residuos sean nulos. Pequeñas diferencias debidas al azar (técnicamente denominadas "no significativas) entre el modelo observado y el modelo esperado (o lo que es lo mismo, pequeños residuos) no deben llevarnos a rechazar la hipótesis nula. El problemas radica en determinar a partir de qué punto esas "pequeñas diferencias" son ya lo suficientemente importantes (o significativas) como para rechazar la hipótesis nula de independencia.

3.- $\alpha = 0'05$

4.- $n = 81$.

5.- valores críticos: σ_x: desconocida, valores "t".



$-t_{(n-1);\alpha/2} = -t_{80;0'025} = -1'9901$

$t_{(n-1);\alpha/2} = t_{80;0'025} = 1'9901$

6.- Estadístico de contraste: $EC = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$

$E.C. = \dfrac{40.800 - 40.000}{5400/9}$; $E.C. = 1'\widehat{33}$

7.- Regla de decisiones:

Si $E.C \geq +t_{n-1;\alpha/2}$ o
Si $EC \leq -t_{n-1;\alpha/2}$ } Rechazo H_0

8.- Aceptamos la hipótesis nula ya que cae dentro de la zona de aceptación.

9.- En el 95% de las ocasiones, los acumuladores Dillco tendrán una duración de 40.000 millas.

10.- En el 5% de las ocasiones nos podemos equivocar aceptando la hipótesis nula, siendo esta falsa. (ERROR TII, con probabilidad $\beta = 0'05$).

3.-

v.a. "X" → "Longitud promedio de una barra".

Nat. X → Normal. (aunque $n < 30$, suponemos que sigue una distr. normal para su resolución).

↓ m.a.s

$n = 25$ $\bar{X} = 2'43$ pies; $\hat{\sigma}_{\bar{X}} = 0'20$ pies.

$\alpha = 0'05$

Debemos contrastar la hipótesis de partida de que la media poblacional sea al menos de 2'8 pies. Para ello realizaremos una prueba estadística y comprobaremos si Ho es verdadera o falsa.

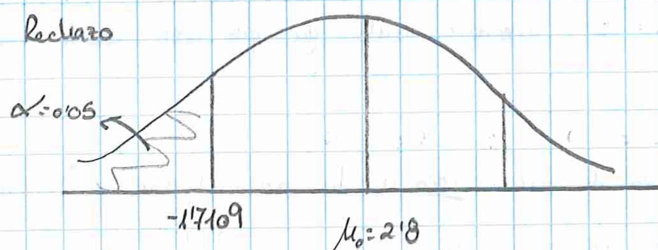
1.- Ho: $\mu_0 \geq 2'8$ pies

2.- Ha: $\mu_0 < 2'8$ pies

3.- $\alpha = 0'05$

4.- $n = 25$

5.- valores críticos: $\hat{\sigma}_{\bar{X}}$: desconocida → valores "t".



$-t_{n-1;\alpha} = -1'7109$

6.- Estadístico de contraste: E.C. = $\bar{X}-\mu_0 / S/\sqrt{n}$

$$E.C. = \frac{2'43 - 2'8}{0'20 / 5} \quad ; \quad E.C. = -9'25$$

7.- Regla de decisión:

Si $E.C \leq -t_{n-1;\alpha}$ Rechazo Ho

8.- Rechazo la Ho porque las diferencias encontradas en la muestra son significativas.

9.- En el 95% de las ocasiones, la longitud de una barra será menor que 2'8 pies.

10.- En el 5% de las ocasiones nos podemos equivocar, rechazando la hipótesis nula siendo esta cierta. (ERROR TI, con una probabilidad $\alpha = 0'05$).

La CIA deberá hacer un ajuste en la máquina, ya que la longitud de la barra en el 95% de las ocasiones es inferior al 2'8 (pies), y la máquina trabaja correcta/ cuando esta longitud es de al menos.

4.-

v.a. "X" → "Promedio de las cuentas por cobrar".

Nat. X → Normal.
↓ m.a.s

$n = 36$ $\bar{X} = 12.300$ \$; $S = 3.300$ \$

μ_0: 10.000 \$

$\alpha = 0'10$.

Debemos contrastar la hipótesis de partida del enunciado, para lo cual formaremos una prueba estadística y vemos si H_0 es verdadera o falsa.

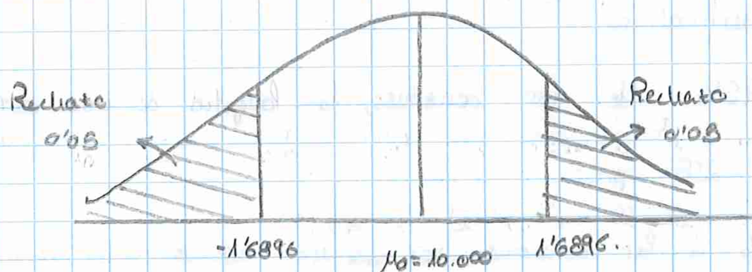
1.- $H_0: \mu_0 = 10.000$ \$

2.- $H_a: \mu_0 \neq 10.000$ \$

3.- $\alpha = 0'1$

4.- $n = 36$

5.- valores críticos:



$t_{n-1;\alpha/2} = t_{35;005} = 1'6896$; $-t_{n-1;\alpha/2} = -t_{35;005} = -1'6896$

6.- Estadístico de contraste: $EC = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$

$E.C. = \frac{12.300 - 10.000}{\frac{3.300}{6}}$; $E.C = 4'1818$

7.- Regla de aceptación:

Si $E.C \leq +t_{n-1;\alpha/2}$
Si $E.C \geq -t_{n-1;\alpha/2}$ } Rechazo H_0.

8.- Rechazo la hipótesis nula pues el "E.C" cae en la zona de rechazo.

9.- En el 90 % de las ocasiones, las cuentas pendientes de cobro no serán de 10.000 \$.

10.- En el 10% de las ocasiones nos podremos equivocar rechazando Ho cuando esta es cierta (ERROR T1 con una probabilidad $\alpha = 0'10$).

5.-

v.a. X -> "Cantidad de tiempo que necesitan los trabajadores para llegar a la C^a"

Nat. X → Normal

↓ m.a.s

$n = 12$; $\bar{X} = 47'9166$, $S = 22'7095$

$\alpha = 0'01$

Debemos contrastar la hipótesis de partida, para lo que realizaremos una prueba estadística y veremos si Ho es verdadera o falsa.

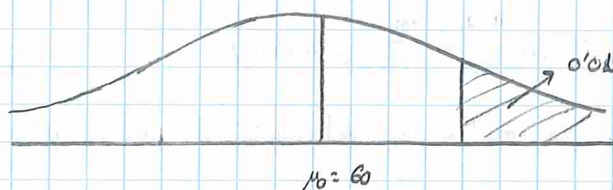
1.- $H_0: \mu_0 \leq 60$

2.- $H_a: \mu_0 > 60$

3.- $\alpha = 0'01$

4.- $n = 12$

5.- valores críticos:



$t_{n-1;\alpha} = t_{11;0'01} = 2'7181$

6.- Estadístico de contraste: $E.C. = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$

$$E.C. = \frac{47'916667 - 60}{\frac{22'7095}{\sqrt{12}}} ; \underline{E.C. = -1'84319}$$

7.- Regla de decisión:

Si $E.C. \geq t_{n-1;\alpha}$ -> Rechazo H_0.

8.- Acepto la hipótesis nula, pues cae en la zona de aceptación.

9.- En el 99% de las ocasiones, los trabajadores emplearán menos de 60 minutos para llegar a su C$^{\underline{a}}$.

10.- En el 1% de los casos puede que nos equivoquemos aceptando la H_0 cuando ésta en realidad es falsa. (ERROR T2, donde $\beta = 0'01$).

6.-

v.a "X" -> "Precio de una calculadora de la marca desconocida".

Nat. X : ¿? TCL -> Normal.

↓ m.a.s

$n = 8$ $\quad \bar{X} \rightarrow N\left(\bar{x} = 9; \ S_x/\sqrt{n} = 0'4629\right)$

v.a "Y" -> "Precio de la calculadora de una marca conocida".

Nat. Y -> ¿? TCL -> Normal.

↓ m.a.s

$n = 8$ $\quad \bar{Y} \rightarrow N(\bar{x} = 11'25; \ S_y/\sqrt{8} = 0'36596)$

Buscamos los límites que acotan la diferencia de medias poblacionales: ¿$\mu_x - \mu_y$?

Las v.a. 'x' e 'y' son dependientes, pues no se puede calcular el precio de una sin el de la otra.

No sabemos el tamaño de la población, pero consideramos que es

lo suficient/ grande para afirmar que $N \to \infty$ y estamos ante MCR.

Definimos una nueva v.a. $d = x - y$.

1.- $H_0: \mu_d \leq 0$ ó $\mu_x - \mu_y \leq 0$.

Unilateral por la dcha.

2.- $H_a: \mu_d > 0$. ó $\mu_x - \mu_y > 0$.

3.- $\alpha = 0'01$.

4.- $n = 8$. (suponemos que siguen una distribución normal).

5.- Estadístico de contraste:

Estadístico muestral $\bar{d}$

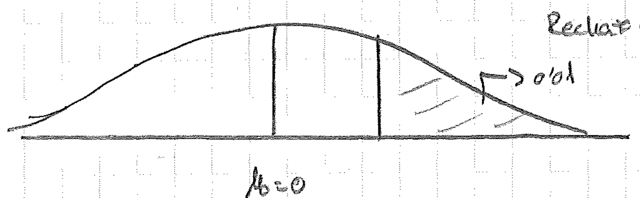
Distribución: $d = x - y \to \bar{d} = -2'25$, $S_d = 1'1649647$.

$\bar{d} \to N\left(\mu_{\bar{d}} = -2'25;\ S_{\bar{d}} = {S_d}/{\sqrt{n}} = 0'41187\right)$

$$E.C. = \frac{\bar{d} - (\mu_x - \mu_y)}{S_d / \sqrt{n}} = t_{n-1}$$

$$E.C. = \frac{-2'25 - 0}{1'1649 / \sqrt{8}} ; \quad \underline{E.C. = -5'4628}$$

6.- Valores críticos:



$t_{n-1;\alpha} = t_{7;0'01} = 2'9980$.

¿Está bien la regla de decisión? ¿P.q. me da lo contrario q. en el enunciado?

7.- Regla de decisión:

Si $E.C \geq t_{n-1;\alpha}$ → Rechazo H_0.

8.- Acepto la hipótesis nula pues cae en la zona de aceptación.

9.- En el 99% de las ocasiones, la marca desconocida se venderá a un precio más bajo.

10.- En el 1% de las ocasiones podemos estar cometiendo un error aceptando la H_0 cuando esta en realidad es falsa.

1.-

v.a. "X" → "Cantidad de bebida que despacha la máquina".

Nat. X ¿?

$\sigma_X = 0'2$

↓ m.a.s

n= 16 → MSR, pero por desconocer N, la consideramos suficiente/ grande como para hacer MCR.

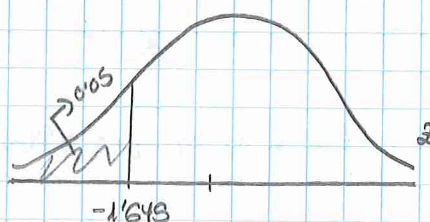
α= 5%

Calculo la probabilidad de un error de tipo II (β) y la potencia de la prueba (1-β) si la cantidad promedio de población despachada es:

1. H_0: $\mu_X \geq 7$
2. H_a: $\mu_X < 7$
3. $\alpha = 0'05$
4. n = 16
5. ¿E.C.?
6. Valor crítico.

v.a. $\bar{X} \to N(\mu_{\bar{X}} : ?\ ;\ \sigma_{\bar{X}}^2 : \sigma_X^2/n = 0'2^2/16)$

$z_{0'95} = -1'645$



CAP III: ESTIMACION DE PARAMETROS.

* ERROR MUESTRAL → Se comete por utilizar muestras para estimar parámetros. (d)

  d = | valor del estimador − valor del parámetro |

* ESTADÍSTICO = ESTIMADOR PUNTUAL:
  - INSESGADO.
  - CONSISTENTE.
  - EFICIENTE.
  - SUFICIENTE.

* NIVEL DE CONFIANZA: Probabilidad de que el intervalo construido en base al estimador puntual incluya al parámetro poblacional desconocido. (1-α)

* ESTIMACION POR INTERVALOS DE LA MEDIA POBLACIONAL:

→ VARIANZA POBLACIONAL CONOCIDA:

Pto. partida: $X \to N(\mu_x; \sigma_x^2)$
$\downarrow$
$\bar{X} \to N(\mu_{\bar{x}} = \mu_x; \sigma_x/\sqrt{n})$

$$Z = \frac{\bar{X} - \mu_x}{\sigma_x/\sqrt{n}}$$

→ $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$

$$P\left(\mu_x \in \bar{X} \pm z_{\alpha/2}\, \sigma_x/\sqrt{n}\right) = 1-\alpha$$

$$P\left(\mu_x \in \bar{X} \pm z_{\alpha/2}\, \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1-\alpha$$ * M.S.R.

→ VARIANZA POBLACIONAL DESCONOCIDA:

Pto. de partida: $X \rightarrow N(\mu_x: ¿?; \sigma_x: ¿?)$

$\downarrow$

$\bar{X} \rightarrow N(\mu_{\bar{x}} = \mu_x;\ S^2/n)$

$$t_{n-1} = \frac{\bar{X} - \mu_x}{S/\sqrt{n}}$$

$$P(-t_{n-1;\alpha/2} \leq t_{n-1} \leq t_{n-1;\alpha/2}) = 1-\alpha$$

$$P\left(\mu_x \in \bar{X} \pm t_{n-1;\alpha/2}\ S/\sqrt{n}\right) = 1-\alpha$$

$$P\left(\mu_x \in \bar{X} \pm t_{n-1;\alpha/2}\ S/\sqrt{n} \sqrt{1-\frac{n}{N}}\right) = 1-\alpha$$

✱ M.S.R.

* CALCULO DEL TAMAÑO DE LA MUESTRA:

a) M.C.R.:

$$d_{máx} = z_{\alpha/2} \frac{\sigma_x}{\sqrt{n}}$$

$\Downarrow$

$$n = z^2_{\alpha/2} \frac{\sigma_x^2}{d^2_{máx}}$$

b) M.S.R.:

$$n_{SR} = \frac{n_{CR}}{1 + \frac{n_{CR}-1}{N}}$$

* Análisis que Involucran a Dos Poblaciones:

a) Varianzas Poblacionales Conocidas:

Pto. de partida: $X \to N(\mu_x; \sigma_x^2)$ $\quad$ $Y \to N(\mu_y; \sigma_y^2)$

$\bar{X} \to N(\mu_x; \sigma_x^2/\sqrt{n})$ $\quad$ $\bar{Y} \to N(\mu_y; \sigma_y^2/\sqrt{n})$

$$\bar{X} - \bar{Y} \to N\left(\mu_x - \mu_y; \sigma_x^2/n_x + \sigma_y^2/n_y\right)$$

CENTRO ± RADIO.

$$(\bar{X} - \bar{Y}) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

b) Varianzas Poblacionales Desconocidas: (Son iguales según prueba de homocedasticidad).

$$\bar{X} - \bar{Y} \to N\left(\mu_x - \mu_y; \frac{{}^*S_p^2}{n_x} + \frac{{}^*S_p^2}{n_y}\right)$$

CENTRO ± RADIO.

$$(\bar{X} - \bar{Y}) \pm t_{n_x+n_y-2, \alpha/2} \sqrt{\frac{S_p^2}{n_x} + \frac{S_p^2}{n_y}}$$

$$* \; S_p^2 = \frac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x + n_y - 2}$$

c) VARIANZAS POBLACIONALES DESCONOCIDAS: (SON DISTINTAS SEGÚN PRUEBA DE HOMOCEDASTICIDAD).

$$(\bar{x}-\bar{y}) \rightarrow N\left(\mu_x-\mu_y;\ \frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}\right)$$

CENTRO ± RADIO

$$(\bar{x}-\bar{y}) \pm t_{k;\alpha/2}\sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}$$

* $$k=\frac{(n_x-1)(n_y-1)}{(n_y-1)c^2+(1-c)^2(n_x-1)}$$

ESTIMACIÓN POR INTERVALOS DE LA PROPORCIÓN POBLACIONAL DE ÉXITOS:

A. CONDICIONES DE APROXIMACIÓN A LA DISTRIBUCIÓN NORMAL:

Pto. de partida: $E(\bar{p})=P$

$Var(\bar{p})=PQ/n$ (M.C.R.)

$=PQ/n \cdot {N-n}/{N-1}$ (M.S.R.)

CENTRO ± RADIO

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

* Para conocer "P" es necesario conocer "PQ". para ello en la variable de partida asignamos los valores "1" y "0" a los éxitos y los fracasos respectivamente. Calculamos así la media y la varianza poblacional (P*Q).

$$\bar{p} \pm z_{\alpha/2}\sqrt{\bar{p}\bar{q}/n-1}$$

$$\bar{p} \pm z_{\alpha/2}\sqrt{\bar{p}\bar{q}/n-1}\sqrt{1-\frac{n}{N}}$$

* M.S.R.

* CÁLCULO DEL TAMAÑO DE MUESTRA NECESARIO PARA ESTIMAR LA PROPORCIÓN POBLACIONAL DE ÉXITOS:

a) M. C. R.:

$$d_{max.} = Z_{\alpha/2} \sqrt{\frac{PQ}{n}}$$

$$\Downarrow$$

$$n = Z^2_{\alpha/2} \frac{PQ}{d^2_{máx.}}$$

b) M. S. R.:

$$n_{SR} = \frac{n_{CR}}{1 + \frac{n_{CR} - 1}{N}}$$

* ESTIMACIÓN POR INTERVALOS DE LA VARIANZA POBLACIONAL:

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right] = 1-\alpha$$

Coste unit. prod. = $\sum$ cant. factores * coste unitario factor.

El coste unit. prod. del período anterior y el actual <u>deben coincidir</u> → Para ello <u>3 condiciones</u>

- Cant. factor por unid. pto. deben mantenerse ctes. (rendimientos).
- El coste unit. factores debe permanecer ctes.
- El nivel de activ. secc. no debe cambiar

↓

Si las condiciones no se cumplen el valor EF en curso estará distorsionado.

Sistemas valoración de la producción.

- **LIFO de factores** → Por cada tipo de pto. una <u>única</u> ficha fabricación. → Para distribuir el coste total de la ficha de fabr. entre unid. terminadas y no terminadas se calcula el coste unit. de prod. del presente período. → El coste unit. así calculado se aplica a todas las unid. terminadas. → Para calcular el valor de la EF en curso:

  Total ficha fabr.
  - Valor unid. ptas acabados

  Valor de la EF en curso →

- **FIFO de factores** → Por cada tipo de pto. tantas fichas fabr. como situaciones diferentes haya. → A cada unid. de pto. le aplicaremos el coste unit. que realmente le corresponde.